desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 1: What Kinds of Data Can I Collect?

Summary

Survey data can be classified as *categorical* or *quantitative*.

	Categorial	**Quantitative**
Example Survey Questions and Data	**Question**: What is your favorite color? *Responses: red, blue, yellow, green* **Question**: Do you usually sleep more than 8 hours a night? *Responses: yes, no, no, yes*	**Question**: How many pairs of shoes do you own? *Responses:* 3, 1, 5, 7 **Question**: How many hours of sleep did you get last night? *Responses:* 8, 9, 7, 8
Your Example Survey Questions and Data	**Question**: ***Responses vary.*** **What is your favorite meal?** *Responses*: ***Responses vary.*** **Breakfast, Lunch, Dinner, Lunch**	**Question**: ***Responses vary.*** **How many seasons does your favorite TV show have?** *Responses*: ***Responses vary.*** 8, 2, 15, 1

In your own words, what is the difference between categorical and quantitative data?

Explanations vary. **Categorical data has values that are categories or words, and quantitative data has values that are numbers, measurements, or quantities.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 1: What Kinds of Data Can I Collect?

Try This!

Decide whether each survey question will produce categorical or quantitative data.

1.1 How many languages do you speak?

Quantitative

1.2 Are you left- or right-handed?

Categorical

1.3 Do you have any pets?

Categorical

1.4 What is your height?

Quantitative

1.5 How many pets do you have?

Quantitative

1.6 Which month were you born in?

Categorical

Write a question that could produce each data set.

2.1 *Responses: swimming, running, walking*

Responses vary. **What is your favorite type of exercise?**

2.2 *Responses:* 10 *min.,* 15 *min.,* 5 *min.*

Responses vary. **How long did it take you to get to school today?**

3.1 Write a question about music that will produce quantitative data.

Responses vary. **How many music concerts have you gone to?**

3.2 Write a question about music that will produce categorical data.

Responses vary. **Who is your favorite musical artist?**

☐ I can explain the difference between quantitative and categorical data.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 2: Revisiting Dot Plots and Histograms

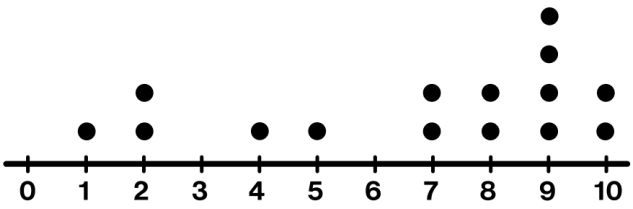
Summary

A *dot plot* and a *histogram* are two ways to visualize quantitative data.

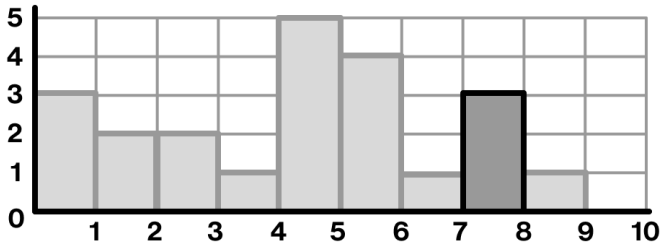
A class played *Love It or Hate It* and rated each season on a scale from 0 to 10.

Here are two representations of their ratings.

Dot Plot of Summer Ratings



Histogram of Winter Ratings



There were 15 ratings for Summer. There were 21 ratings for Winter.

The highest rating for Summer was 10. For Winter, it was between 8 and 9.

A new student gave winter a 7. 7. Add this data point to the histogram above.

What are some advantages of representing data with a histogram? A dot plot?

Explanations vary. **One advantage of representing data with a histogram is that you can use bins, which make it easier to organize data that includes decimals. One advantage of representing data on a dot plot is that you can see all of the individual data points.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 2: Revisiting Dot Plots and Histograms

Try This!

Here is a histogram of students' ratings for the fall season.

Decide if each statement is true, false, or cannot be determined.

1.1 There are 29 total ratings.

True | False | Cannot be determined

1.2 The highest rating included was a 9.9.

True | False | **Cannot be determined**

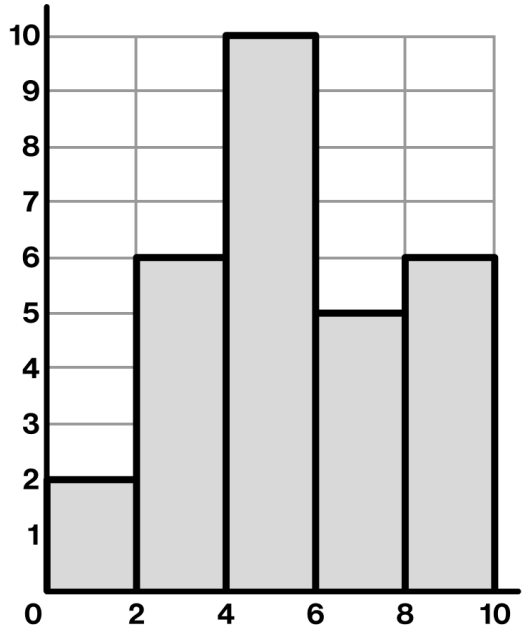
1.3 The lowest rating was less than 2.

True | False | Cannot be determined

1.4 There are 10 ratings higher than 6.

True | **False** | Cannot be determined

Histogram of Fall Ratings



2. Here are students' ratings for spring: $4.5,\ 5.1,\ 5.6,\ 6.5,\ 6.9,\ 7.1,\ 7.4,\ 7.9,\ 8.4$.

Why might someone make a histogram over a dot plot to visualize this data set?

Explanations vary. **Someone might make a histogram over a dot plot because a histogram can allow us to organize decimals in bins that can be easier to see. For example, grouping all of the ratings between 7 and 8.**

- ☐ I can use technology to represent data with a dot plot or histogram.
- ☐ I can describe the advantages and disadvantages of using a dot plot or a histogram to represent data.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 3: Revisiting Box Plots

Summary

A *box plot* can be used to visualize a one-variable quantitative data set.

Zahra used a fitness app to track how many miles she walked on foot. Here is a box plot of daily miles traveled on foot each day by Zahra in June.

Complete the definitions and identify the statistics for Zahra's data.

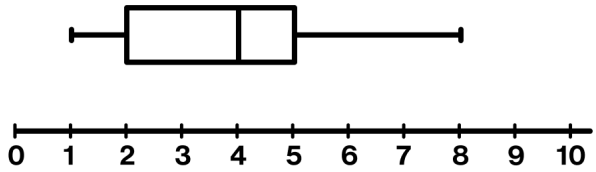
Minimum: The smallest value

Quartile 1: The middle of the lower half of the data

Median: **The middle of the entire data set**

Quartile 3: **The middle of the upper half of the data**

Maximum: **The largest value**



Min.	Q1	Median	Q3	Max.
1	2	4	5	8

Select **all** the statements that are true according to the box plot.

- ☐ Zahra's mean miles walked in June was 5 miles.
- ☐ The middle 50% of miles walked were between 1 and 8.
- ✓ **Zahra never walked 9 miles in June.**
- ☐ There was one day Zahra walked 3 miles**.**
- ✓ **Zahra walked 4 miles or less for half of the days in June.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

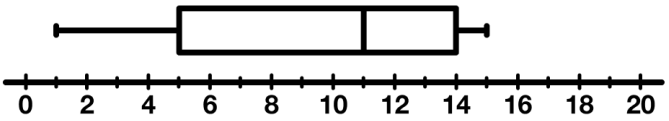
Lesson 3: Revisiting Box Plots

Try This!

Two basketball players recorded their points for each game in the season.

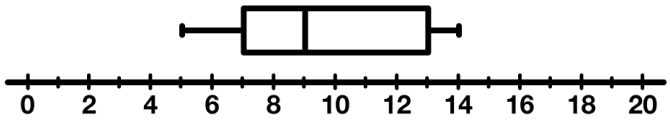
Use the box plots of their data to identify each statistic.

1.1 Basketball Player A



Min.	Q1	Median	Q3	Max.
1	5	11	14	15

1.2 Basketball Player B



Minimum	Median	Maximum
5	9	14

Decide if each statement is true, false, or cannot be determined.

2.1 Player A played 15 games this season.

True False **Cannot be determined**

2.2 In half of Player B's games, they scored 9 points or fewer.

True False Cannot be determined

2.3 Player A scored 13 points in at least one game.

True False **Cannot be determined**

2.4 Player A scored 0 points in a game.

True **False** Cannot be determined

- ☐ I can interpret the parts of a box plot and use technology to represent data with a box plot.
- ☐ I can use box plots to compare data sets.

desmos

Unit A1.3, Describing Data: Notes

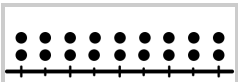
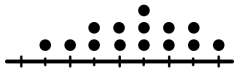
Answer Key

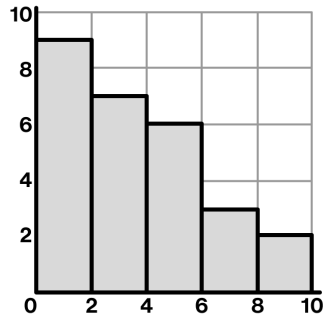
Lesson 4: Describing Data Sets

Summary

The shapes of data can be described as *bimodal*, *uniform*, *symmetric*, *skewed*, and *bell-shaped*.

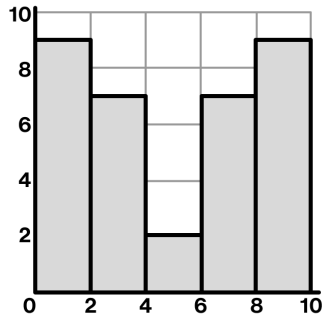
Create the missing definitions or sketches.

Shape Description	Dot Plot	Definition
Bimodal		**There are two peaks in the data.**
Uniform		Data values are evenly distributed.
Symmetric		**The data has a line of symmetry.**
Skewed		One side of the data has more values than the other.
Bell-Shaped		**Most of the data is at the center with fewer points farther from the center.**



Shape Description:

Skewed



Shape Description:

Bimodal and Symmetric

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

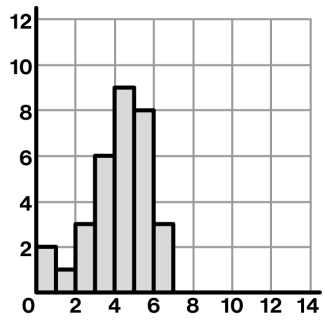
Answer Key

Lesson 4: Describing Data Sets

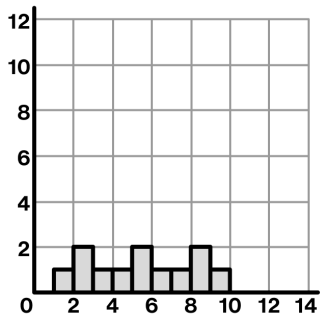
Try This!

Match each histogram with the best description of its shape.

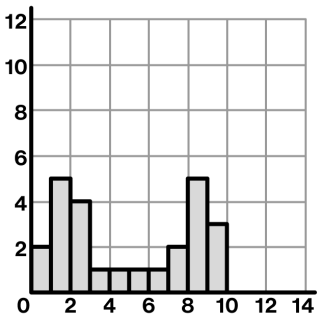
1.1 **C) Skewed**



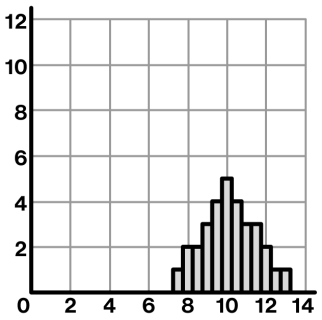
1.2 **D) Symmetric**



1.3 **A) Bimodal**



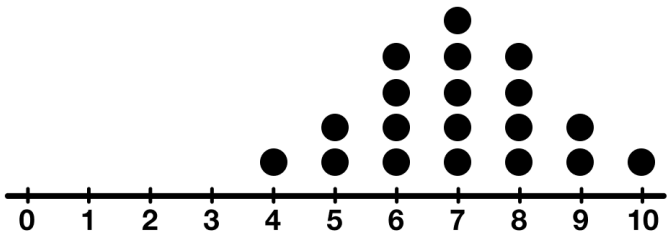
1.4 **B) Bell-shaped**



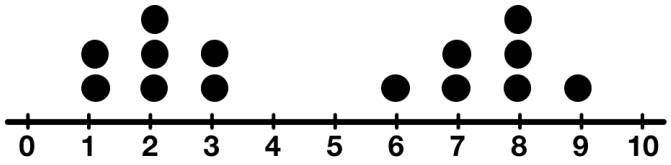
A. Bimodal	B. Bell-shaped	C. Skewed	D. Symmetric

Sketch a dot plot or histogram that matches each description.

2.1 Bell-shaped



2.2 Bimodal



☐ I can describe the shape of data sets represented with dot plots, histograms, and box plots.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 5: Revisiting Measures of Center

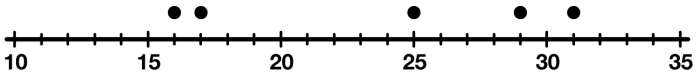
Summary

Mean and *median* are two *measures of center* used to describe data sets.

The shape of the data can influence which measure of center to use.

Here is a dot plot of Oscar's scores from a video game. Calculate the mean and median.
Use the Unit 3 Calculator Guide if it helps with your thinking.

Mean	Median
23.6	25



Here is a histogram of starting salaries (in thousands of dollars) at Des-Cafe.

Mean	Median
33.5	29.5

What is the shape of the data? **Skewed**

Explain why someone might say the *median* is more representative of a typical starting salary.

Explanations vary. **Someone might say the median is more representative of the typical starting salary because the shape is skewed. The higher values above 40 thousand have a big impact on the mean compared to the median.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 5: Revisiting Measures of Center

Try This!

Use the Desmos Graphing Calculator to create a dot plot or histogram of each data set and calculate the mean and median. Use the Unit 3 Calculator Guide if it helps with your thinking.

1.1 DesWash n' Go hourly wages (in dollars)

12	13	13
14	14	14
15	15	16

Mean: 14

Median: 14

Which is larger? **They are the same**

Shape: **Symmetric and bell-shaped**

1.2 DesTunes Music hourly wages (in dollars)

12	12	13
13	13	15
17	18	19

Mean: 14.67

Median: 13

Which is larger? **Mean**

Shape: **Skewed**

The worker making $16 an hour is promoted to $22 an hour. Which measure would increase?

Circle One: **mean** / median / both / neither

Explain your thinking.

Explanations vary. **The mean will increase because one of the values increased. The median will stay the same because changing the maximum value in this data set does not affect the middle of the data set.**

A new worker is hired and will make $20 an hour. Which measure would increase?

Circle One: mean / median / **both** / neither

Explain your thinking.

Explanations vary. **The mean would increase because adding a value higher than the mean will increase the mean. The median will increase because we are including a new wage that is changing the middle of the data set.**

- ☐ I can explain how to calculate the mean and median and what these tell us about a data set.
- ☐ I can use technology to calculate the measure of centers (mean and median) for a data set.
- ☐ I can explain the effect of extreme values on the mean and median.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

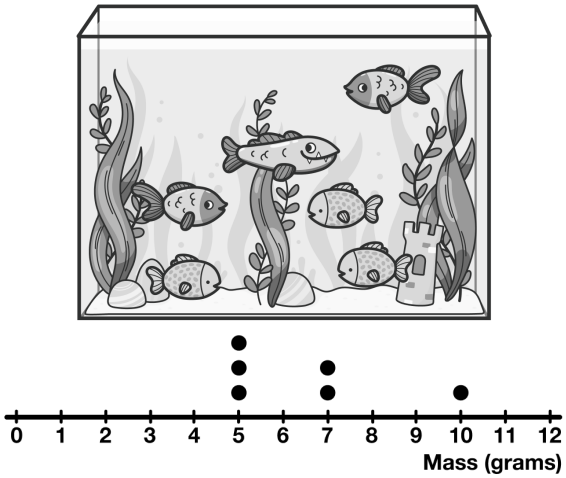
Lesson 6: Introduction to Standard Deviation

Summary

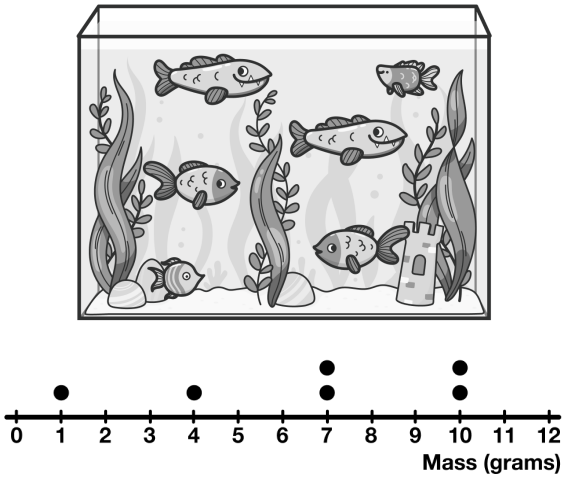
One way to measure the consistency or *spread of data* is to calculate its *standard deviation*. Data with a larger standard deviation is more spread out than data with a smaller standard deviation.

Here are the masses (in grams) of the fish in two new tanks. Calculate the statistics for Tank B.

Tank A: $5, 5, 5, 7, 7, 10$



Tank B: $1, 4, 7, 7, 10, 10$



Mean	Standard Deviation
$A = [5, 5, 5, 7, 7, 10]$ $\text{mean}(A)$ $= 6.5$	$A = [5, 5, 5, 7, 7, 10]$ $\text{stdevp}(A)$ ≈ 1.8

Mean	Standard Deviation
$B = [1, 4, 7, 7, 10, 10]$ $\text{mean}(B)$ $= 6.5$	$B = [1, 4, 7, 7, 10, 10]$ $\text{stdevp}(B)$ ≈ 3.2

Describe what the mean and standard deviation say about how the fish in Tanks A and B compare.

Explanations vary. **On average, fish in Tank A will weigh about the same as fish in Tank B since their means are the same. Tank A's fish have a more consistent weight than Tank B's fish because Tank A has a smaller standard deviation.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

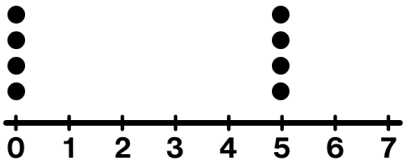
Lesson 6: Introduction to Standard Deviation

Try This!

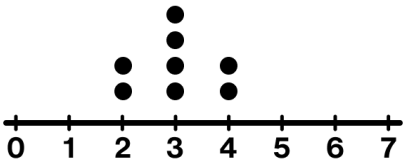
1.1 Which dot plot do you think has the greatest standard deviation? **A**

1.2 Which dot plot do you think has the lowest standard deviation? **B**

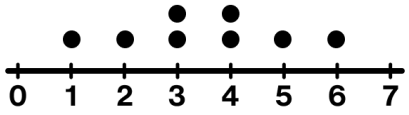
A.



B.



C.



Calculate the mean and standard deviation for each of the data sets above.
Use a calculator to help you with your thinking.

2.1 Data Set A

Mean	Standard Deviation
2.5	2.5

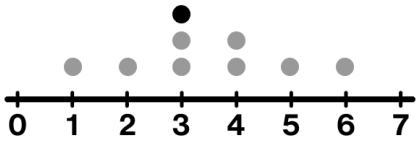
2.2 Data Set B

Mean	Standard Deviation
3	0.71

2.3 Data Set C

Mean	Standard Deviation
3.5	1.5

3. Add a data point to this dot plot that will lower the standard deviation.



Explain your thinking. ***Explanations vary.***

I added a data point at 3 to lower the standard deviation because I can tell from this graph that the mean is about 3.5. When I add a point near the mean, it lowers the standard deviation.

- ☐ I understand that standard deviation is a measure of spread and can use it to compare data sets.
- ☐ I can use technology to calculate the standard deviation of a data set.

desmos

Unit A1.3, Describing Data: Notes **Answer Key**

Lesson 7: Comparing Data Using Mean and Standard Deviation

Summary

Mean and standard deviation can be used to compare the center and spread of data sets.

Here are the high temperatures (in degrees Fahrenheit) in Desmopolis and Destown last week.

Desmopolis: $65, 73, 80, 82, 79, 68, 71$ Destown: $70, 75, 76, 74, 77, 75, 74$

Calculate the statistics for Destown. Use the Unit 3 Calculator Guide if it helps with your thinking.

City	Mean (°F)	Standard Deviation (°F)
Desmopolis	74	6
Destown	74.43	2.06

What does the mean help compare about the temperatures in different places?

The mean compares which city has higher temperatures in general.

What does the standard deviation help compare about the temperatures in different places?

The standard deviation compares which city has more consistent temperatures.

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 7: Comparing Data Using Mean and Standard Deviation

Try This!

Calculate the mean and standard deviation of the commute times (in minutes) for each traveler.

1.1 Traveler A: $31, 25, 28, 34, 31, 29, 30$

Mean	Standard Deviation
29.71	2.60

1.2 Traveler B: $36, 36, 41, 40, 43, 41, 34$

Mean	Standard Deviation
38.71	3.10

1.3 Traveler C: $30, 29, 38, 42, 47, 45, 44$

Mean	Standard Deviation
39.29	6.71

1.4 Traveler D: $41, 38, 30, 31, 20, 18, 19$

Mean	Standard Deviation
28.14	8.68

1.5 Order the travelers' commute times from least consistent to most consistent.

Least Consistent Traveler D Traveler C Traveler B Traveler A **Most Consistent**

1.6 Which traveler had the longest commute time? Use statistics to justify your thinking.

Responses vary. **I think Traveler C has the longest commute time because their mean travel time is the longest. They also have the single longest travel time compared to all of the other travelers.**

☐ I can use the mean and standard deviation to compare two data sets.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 8: Comparing Data Using Median and IQR

Summary

The *interquartile range* (or *IQR*) measures the middle half of a data set, or the distance between the first and third quartiles.

Here are box plots of the distances traveled by three racecars. Identify the statistics for each car.

Car A

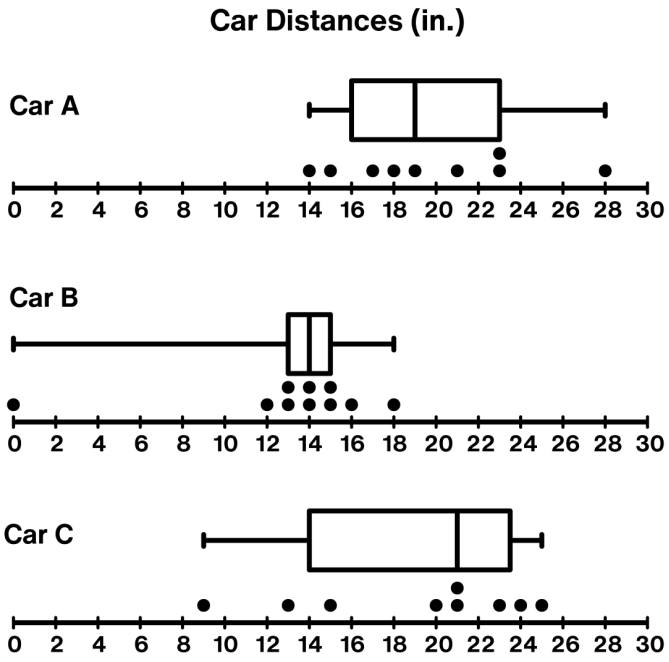
Q1	Q3	IQR	Median
16	23	7	19

Car B

Q1	Q3	IQR	Median
13	15	2	14

Car C

Q1	Q3	IQR	Median
14	23.5	9.5	21



Which car is the most consistent? ***Responses vary.*** **Car B**

Explain which statistics you used to decide.

Responses vary. **Car B is the most consistent because the IQR is the smallest.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

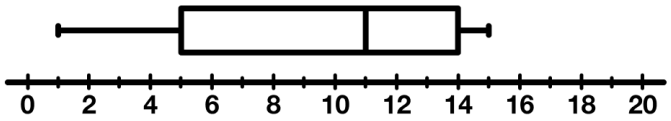
Lesson 8: Comparing Data Using Median and IQR

Try This!

Two basketball players recorded their points for each game in the season.

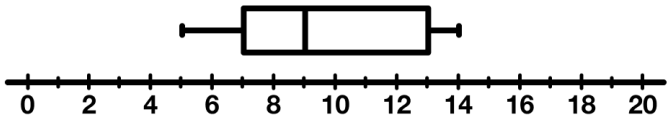
Use the box plots of their data to identify each statistic.

1.1 Basketball Player A



Q1	Q3	IQR	Median
5	14	9	11

1.2 Basketball Player B



Q1	Q3	IQR	Median
7	13	6	9

2.1 Which player was more consistent in their points scored? Explain how you know.

Explanations vary. **Player B was more consistent because their IQR was smaller.**

2.2 Which player generally scored more points? Explain how you know.

Explanations vary. **Player A generally scored more points because their median was higher.**

- ☐ I can calculate the IQR of a data set and understand that it is a measure of spread.
- ☐ I can use medians and IQRs to compare skewed data sets.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

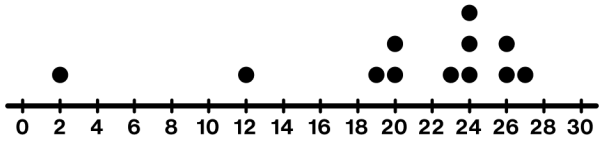
Lesson 9: Identifying Outliers

Summary

Data points that are far from other values in a data set are called *outliers*.

Here are Koharu's scores from a different game.

The mean is 20.58 and the median is 23.5.



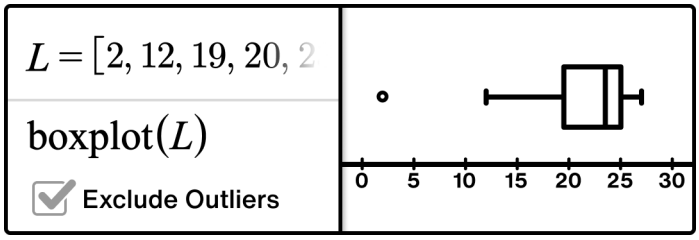
Do you think there are any outliers? Why or why not?

Explanations vary. **Yes, because the data point of 2 is far from the rest of the points in the data set.**

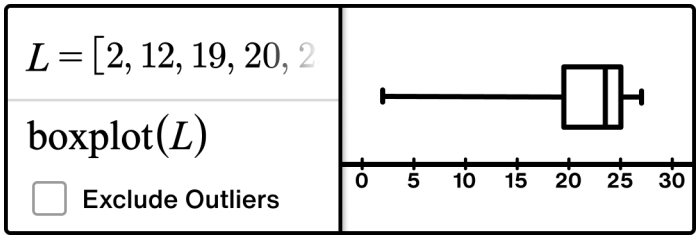
A box plot can help confirm whether or not values in a data set are outliers.

1. Enter the data as a list in the Desmos Calculator.
2. Create a box plot. Select “Exclude Outliers” to see each outlier as its own point.

Box Plot With Outliers Excluded



Box Plot With Outliers Included



Are there any outliers in Koharu’s data? Explain your thinking.

Explanations vary. **Yes, at 2 because the box plot shows an open circle, which represents an outlier.**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

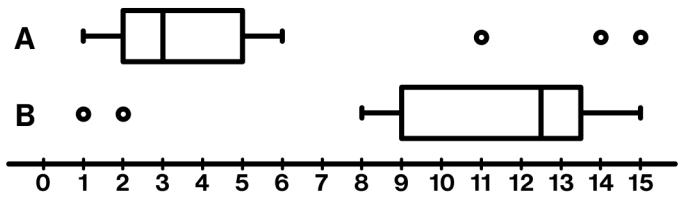
Lesson 9: Identifying Outliers

Try This!

Use the box plot to identify any outliers in each data set.

1. Data Set A outliers: $11,\ 14,\ 15$

2. Data Set B outliers: $1,\ 2$



Here are dot plots that show the number of strikeouts thrown by two pitchers.

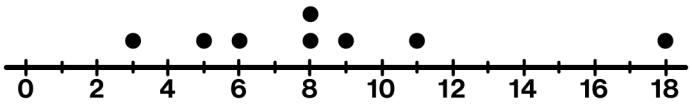
Use a calculator to make a box plot and identify any outliers in each data set.

Use the Unit 3 Calculator Guide if it helps with your thinking.

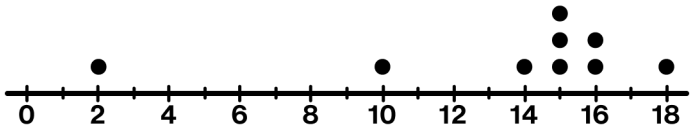
2.1 Pitcher A outliers: 18

2.2 Pitcher B outliers: 2

Pitcher A



Pitcher B



- ☐ I can determine whether or not a data point is an outlier.
- ☐ I can explain how outliers impact the mean or median of a data set.

desmos

Unit A1.3, Describing Data: Notes

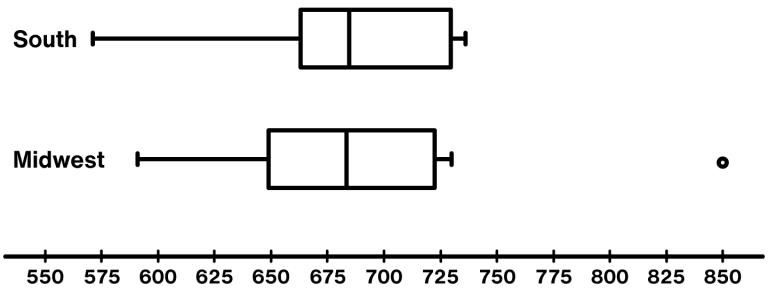
Answer Key

Lesson 10: Comparing Data Using Measures of Center and Spread

Summary

Measures of center (mean/median) and spread (standard deviation/IQR) can help us make sense of and compare data sets. The shape of the data can influence which statistics to use.

Here is the median monthly rent (in dollars) of eight states from the South and Midwest in 2019.



Median Monthly Rent: South

$731, $736, $728, $667
$571, $659, $672, $697

Median Monthly Rent: Midwest

$591, $669, $682, $848
$715, $685, $629, $730

Complete the table with statistics about rents in the Midwest using the Desmos calculator.

Region	Mean	Standard Deviation	Median	IQR	Outliers
South	$682.63	$51.03	$684.50	$66.50	none
Midwest	$693.63	$71.71	$683.50	$73.50	$848

What measures of center and spread would you use to compare the rents in each group of states? Explain your thinking.

Responses vary. **I would use the median and IQR to compare the rents because some of the rents are really low and high in each region, including an outlier at $848 which affects the mean and standard deviation.**

How do the rents in the Midwest compare to the rents in the South?
Use the measures of center and spread you chose above to explain your thinking.

Explanations vary. **The median rent for both regions are alike. The rents in the South are more consistent since the South rents have a smaller IQR of $66.50 compared to the Midwest's IQR of $73.50**

Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

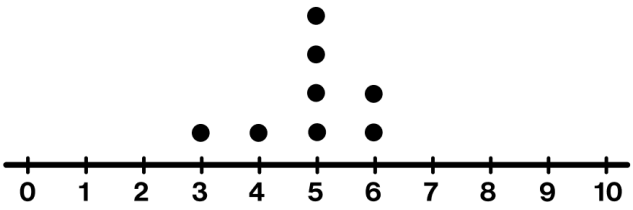
Answer Key

Lesson 10: Comparing Data Using Measures of Center and Spread

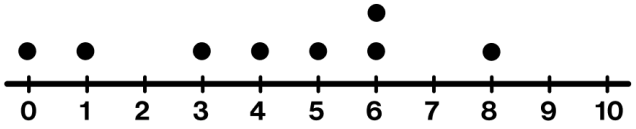
Try This!

Here are the number of free throws that Eva and Camila made during the playoffs.

Eva: 3, 5, 5, 6, 4, 5, 5, 6



Camila: 0, 1, 3, 4, 5, 6, 6, 8



1. Complete the table with the statistics for Camila.
   Use the Unit 3 Calculator Guide if it helps with your thinking.

Player	Mean	Standard Deviation	Median	IQR	Outliers
Eva	4.88	0.93	5	1	none
Camila	4.13	2.52	4.5	4	none

2. Would you use mean or median to compare Eva's and Camila's number of free throws?

   Explanations vary. **I would use the median because Camila's data is skewed.**

3. Compare Eva's and Camila's number of free throws.
   Use statistics about center and spread to support your ideas.

   Explanations vary. **Eva typically scored more free throws than Camila because she had a higher median than Camila. Eva was a more consistent free throw shooter because she had a smaller IQR and standard deviation than Camila.**

☐ I can use statistics appropriate to the shape of the data to compare two data sets.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

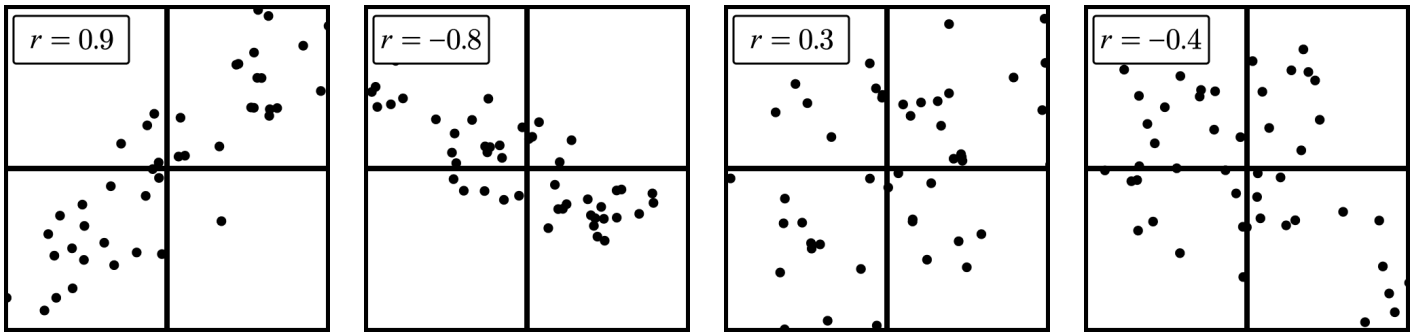
Lessons 11–12: Interpreting Correlation Coefficient in Context

Summary

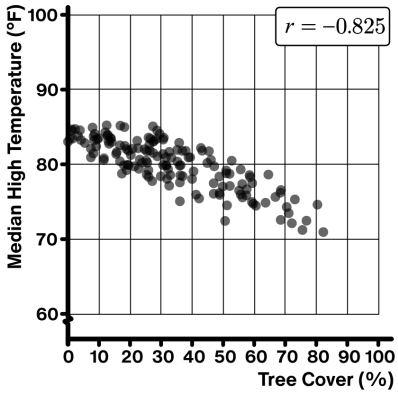
When the points on a scatter plot follow a line, we say there is a *linear association* between x and y.

The r*-value*, also called the *correlation coefficient*, describes the strength (weak, strong) and direction (negative, positive) of an association.

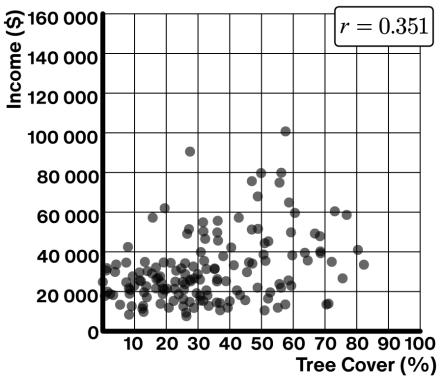
Strong and Positive | **Strong** and **Negative** | **Weak** and **Positive** | **Weak** and **Negative**



Here are two scatter plots with data recorded for 150 blocks in Detroit, Michigan.



Description:
The r-value is -0.825. This means there is a negative and strong relationship between tree cover $\%$ and median high temperature in Detroit, Michigan.



Description:
The r-value is 0.351. This means…
there is a positive and weak relationship between tree cover $\%$ and income in Detroit, Michigan.

Things I Want to Remember

desmos

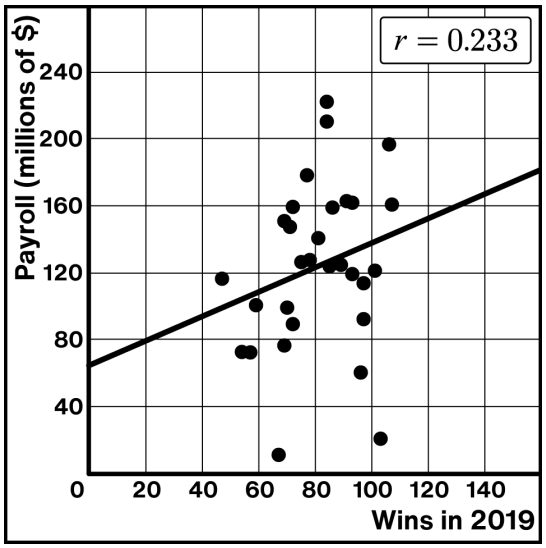
Unit A1.3, Describing Data: Notes

Answer Key

Lessons 11–12: Interpreting Correlation Coefficient in Context

Try This!

Use the correlation coefficient to describe the association shown in each scatter plot.

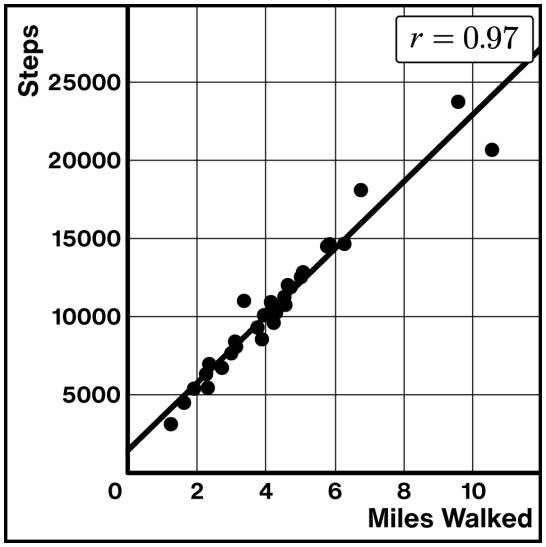


1.1 Lucy was curious about the relationship between money and wins in baseball.

She found data about:

- Payroll (in millions of dollars)
- Wins in 2019

Description: The r-value is 0.233. This means . . . **there is a positive and weak relationship between wins in 2019 and payroll in millions of dollars.**



1.2 Daeja tracks her fitness data on her watch.

She recorded data about:

- Steps
- Miles walked

Description: The r-value is 0.97. This means . . . **there is a positive and strong relationship between Daeja's miles walked and steps.**

- ☐ I can use a correlation coefficient to describe the strength and sign of the relationship between variables on a scatter plot.
- ☐ I can use technology to calculate the correlation coefficient of data on a scatter plot.
- ☐ I can use a correlation coefficient to describe the strength and direction of a linear association.
- ☐ I can interpret correlation coefficients in context.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

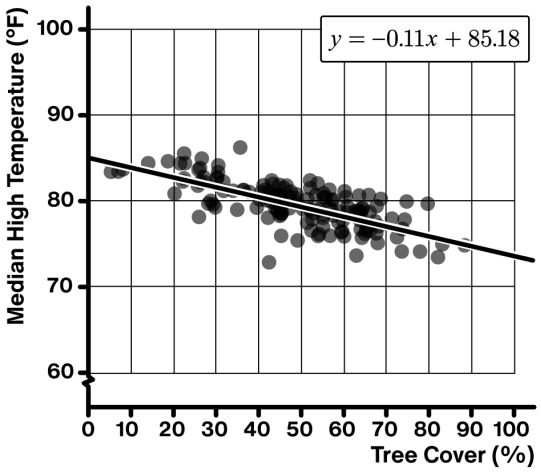
Lesson 13: Interpreting Slope and Vertical Intercept in Context

Summary

Mathematicians use lines of fit to describe linear associations and make predictions.

Here are the median high temperatures and tree covers (%) for 150 blocks in two different cities.

St. Paul, Minnesota

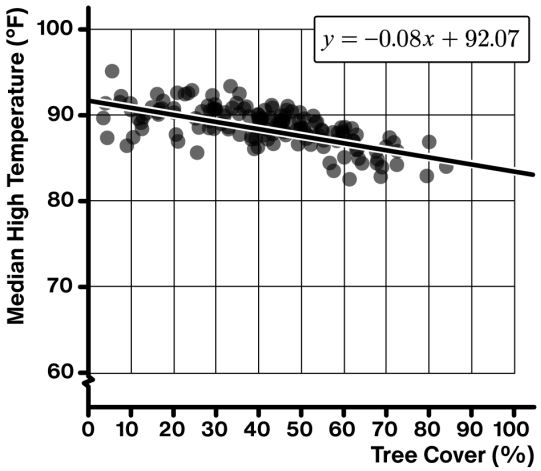


Slope interpretation:
When the tree cover increases by 1% in St. Paul, the predicted temperature decreases by 0.11°F.

y-intercept interpretation:
If the tree cover in St. Paul is 0%, the predicted temperature is 85.18°F.

Prediction: If a block in St. Paul has 80% tree cover, the predicted median high temperature will be about 75°F.

Austin, Texas



Slope interpretation: ***Explanations vary.***

When the tree cover increases by 1% in Austin, the predicted temperature decreases by 0.08°F.

y-intercept interpretation: ***Explanations vary.***

If the tree cover in Austin is 0%, the predicted temperature is 92.07°F.

Prediction: If a block in Austin has 80% tree cover, . . .
the predicted temperature will be about 85°F.

Things I Want to Remember

desmos

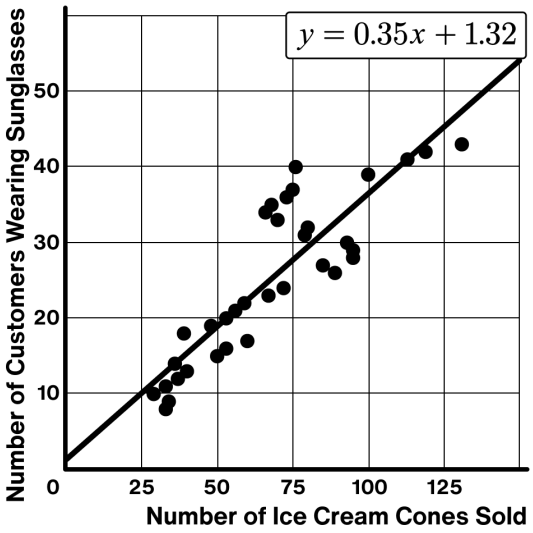
Unit A1.3, Describing Data: Notes

Answer Key

Lesson 13: Interpreting Slope and Vertical Intercept in Context

Try This!

1.1 Nyanna noticed a trend at an ice cream shop. She recorded the number of ice cream cones sold and the customers wearing sunglasses one day.



Slope interpretation: ***Explanations vary.***

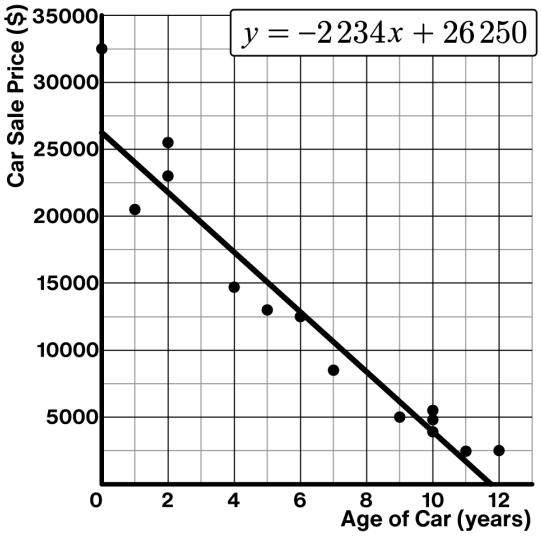
When the number of ice cream cones sold increases by 1 **cone, the predicted number of customers wearing sunglasses increases by** 0.35 **customers.**

y**-intercept interpretation**: ***Explanations vary.***

If the number of ice cream cones sold is 0**, the predicted number of customers wearing sunglasses would be** 1.32 **customers.**

Prediction: ***Predictions vary.*** If 30 cones are sold, **the predicted number of customers wearing sunglasses would be about** 12 **customers.**

1.2 Kwasi was curious about the relationship between the ages of cars and their values. He found data on the ages of several cars (in years) and their sale prices (in dollars).



Slope interpretation: ***Explanations vary.***

When the age of the car increases by 1 **year, the predicted car sale price decreases by** $\$2234$**.**

y**-intercept interpretation**: ***Explanations vary.***

If the age of the car is 0 **years old, the predicted car sale price is** $\$26\,250$**.**

Prediction: ***Predictions vary.*** If a car is 3 years old, . . . **the predicted sale price will be about** $\$20\,000$.

- ☐ I can describe the slope and vertical intercept for a linear model in everyday language.
- ☐ I can estimate unknown values using a line of fit on a graph.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 14: Residuals and Residual Plots

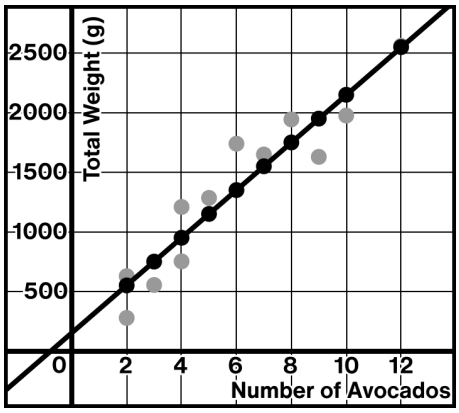
Summary

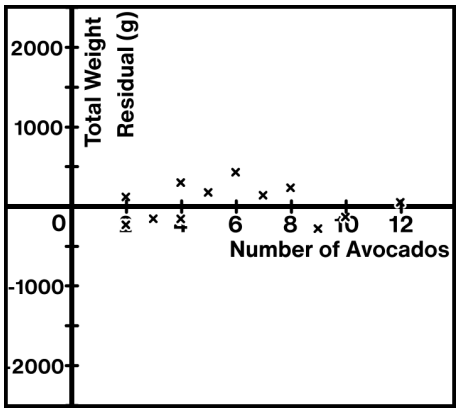
A *residual* is the difference between the y-value of a data point and the value predicted by the line of best fit. A scatter plot of all the residuals (a *residual plot*) can help us decide if a line fits the data well.

On the left is data and a line of fit for several orders of avocados. On the right is its residual plot.

Use the residual plot to explain how you know this line is a good fit for the data.

Explanations vary. **The line is a good fit for the data because the residuals are close to the x-axis and there are random points below and above the x-axis.**

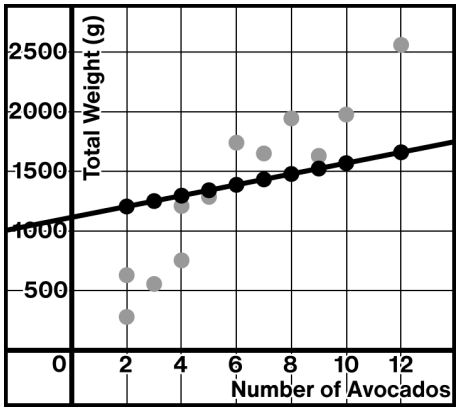


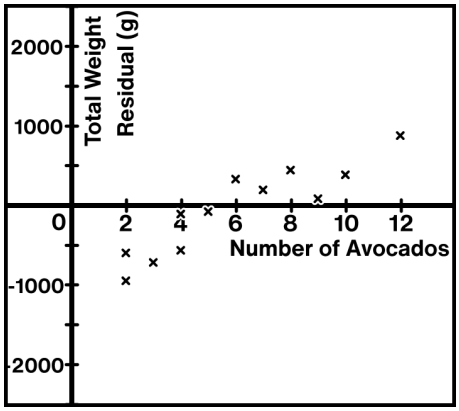


Here is a different line of fit for the data and its residual plot.

Use the residual plot to explain how you know this line is **not** a good fit for the data.

Explanations vary. **The line is not a good fit for the data because many of the points are far from the x-axis. The residuals start off all negative and then turn positive, which shows that the line does not follow the pattern of the data.**





Things I Want to Remember

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lesson 14: Residuals and Residual Plots

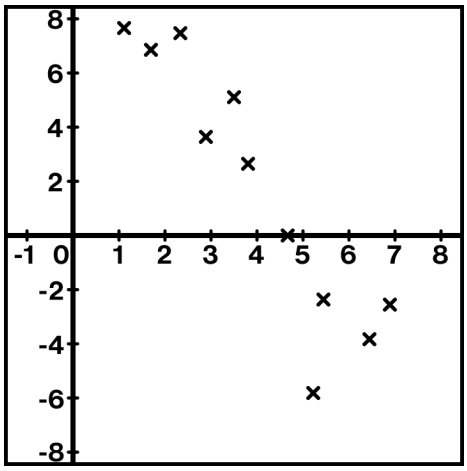
Try This!

1. Describe what the residual plot for a good line of fit looks like.

   Explanations vary. **The residual plot for a good line of fit will have point values that are close to the x-axis and will have random values above and below the x-axis.**

Here are residual plots for lines that are not shown. Describe how you think each line fits the data.

2.1

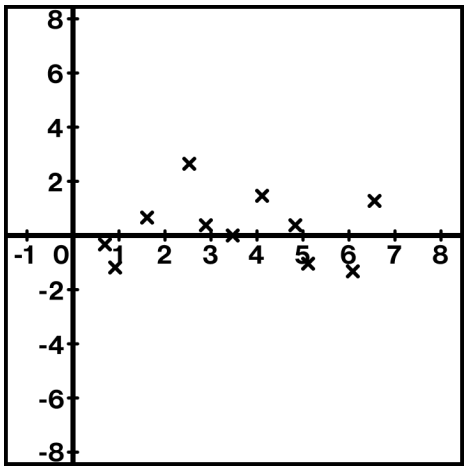


Circle one:
The line fits the data well / **not well.**

Explain your thinking.

Explanations vary. **The residual plot has points that are very far from the x-axis. Also, the residual values go from positive to negative, and are not random.**

2.2



Circle one:
The line fits the data **well** / not well.

Explain your thinking.

Explanations vary. **The residual plot has points that are close to the x-axis and has random points above and below the x-axis.**

- ☐ I can make connections between a residual plot and residuals on a graph.
- ☐ I can recognize when a residual plot indicates a better or worse fit.

desmos

Unit A1.3, Describing Data: Notes

Answer Key

Lessons 15–17: Using Technology to Analyze Two-Variable Data

Summary

A calculator can compute the *line of best fit* and the correlation coefficient to help describe the relationship (or correlation) between two variables. *Causation* is one type of *correlation*.

In a causal relationship, a change in one variable causes a change in the other variable.

Nyanna noticed a trend at an ice cream shop. She recorded the number of ice cream cones sold and the customers wearing sunglasses one day.

Nyanna used a calculator to generate a line of best fit.

Line of best fit equation:

$$y = 0.35x + 1.32$$

The r-value is 0.87. This means . . .

Explanations vary. **There is a positive and strong relationship between the number of ice cream cones and the number of customers wearing sunglasses.**

Use Nyanna's model to predict the number of ice cream cones sold if there are 150 people wearing sunglasses.

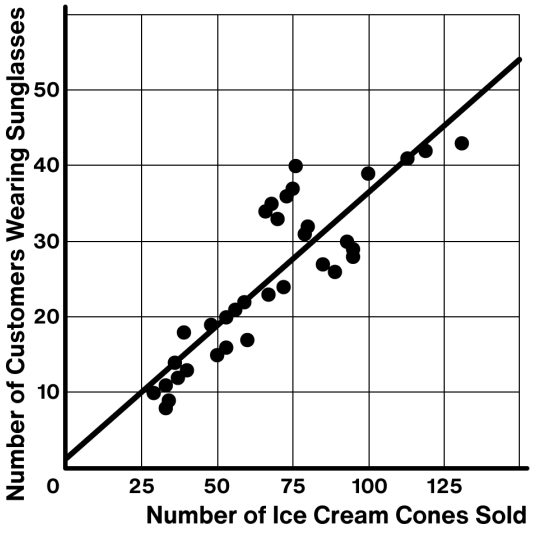
$y = 0.35(150) + 1.32$

$y = 53.82$

~ 54 **ice cream cone sold**

Do you think one of the variables causes the other? If not, what else could be affecting the relationship?

Responses vary. **I do not think one of the variables causes the other. If it is sunny out, people might be more likely to wear sunglasses and to buy ice cream.**



$y_1 \sim mx_1 + b$

STATISTICS	PARAMETERS
$r^2 = 0.7642$	$m = 0.351312$
$r = 0.8742$	$b = 1.31984$

Things I Want to Remember

desmos

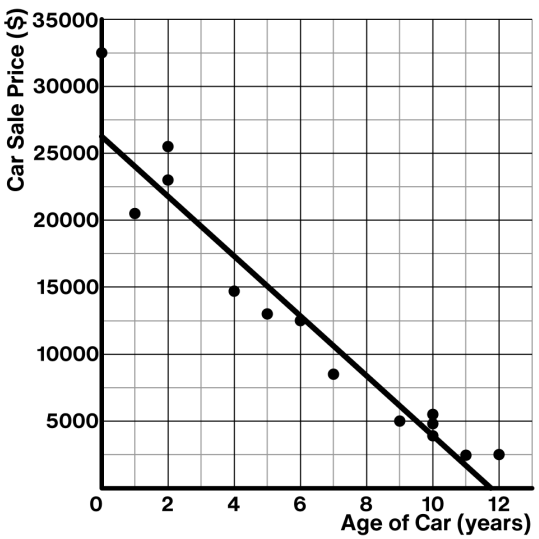
Unit A1.3, Describing Data: Notes

Answer Key

Lessons 15–17: Using Technology to Analyze Two-Variable Data

Try This!

Kwasi was curious about the relationship between the ages of cars and their values. He found data on the ages of several cars (in years) and their sale prices (in dollars).



$y_1 \sim mx_1 + b$

STATISTICS	PARAMETERS
$r^2 = 0.9215$	$m = -2270.38$
$r = -0.96$	$b = 26886.7$

1. **Line of best fit equation:**

   $y = -2\,270.38x + 26\,886.7$

2. The r-value is -0.96. This means . . .

   Explanations vary. **There is a negative and strong relationship between the age of the car and the car sale price.**

3. What does the model predict the price would be for a car that was 8 years old?

   $y = -2\,270.38(8) + 26\,886.7$

   $y = 8\,723.66$

   The predicted price will be $\$8\,723.66$

4. Do you think one of the variables causes the other?

   If not, what else could be affecting the relationship? Explain your thinking.

   Explanations vary. **Yes. I believe that the age of the car causes the price to go down because an older car is more likely to have more mileage or other mechanical issues.**

- ☐ I can use technology to generate the line of best fit for data on a scatter plot.
- ☐ I can use the equation of the best fit line to make predictions.
- ☐ I can determine if the relationship between two variables represents correlation or causation.
- ☐ I can analyze the relationship between two variables in context using scatter plots, lines of best fit, and correlation coefficients.